

A lower bound for metric 1-median selection *

Ching-Lueh Chang ^{†‡}

January 13, 2014

Abstract

Consider the problem of finding a point in an n -point metric space with the minimum average distance to all points. We show that this problem has no deterministic $o(n^2)$ -query $(4 - \Omega(1))$ -approximation algorithms.

1 Introduction

Given oracle access to a metric space $(\{1, 2, \dots, n\}, d)$, the METRIC 1-MEDIAN problem asks for a point with the minimum average distance to all points. Indyk [8, 9] shows that METRIC 1-MEDIAN has a Monte-Carlo $O(n/\epsilon^2)$ -time $(1 + \epsilon)$ -approximation algorithm with an $\Omega(1)$ probability of success. The more general METRIC k -MEDIAN problem asks for $x_1, x_2, \dots, x_k \in \{1, 2, \dots, n\}$ minimizing $\sum_{x \in \{1, 2, \dots, n\}} \min_{i=1}^k d(x_i, x)$. Randomized as well as evasive algorithms are well-studied for METRIC k -MEDIAN and the related k -means problem [1, 4, 7, 10–12], where $k \geq 1$ is part of the input rather than a constant.

This paper focuses on *deterministic sublinear-query* algorithms for METRIC 1-MEDIAN. Guha et al. [7, Sec. 3.1–3.2] prove that METRIC k -MEDIAN has a deterministic $O(n^{1+\epsilon})$ -time $O(n^\epsilon)$ -space $2^{O(1/\epsilon)}$ -approximation algorithm that reads distances in a single pass, where $\epsilon > 0$. Chang [3] presents a deterministic non-adaptive $O(n^{1.5})$ -time 4-approximation algorithm for METRIC 1-MEDIAN. Wu [14] generalizes Chang’s result by showing an $O(n^{1+1/h})$ -time $2h$ -approximation algorithm for any integer $h \geq 2$. On the negative side, Chang [2] shows that METRIC 1-MEDIAN has no deterministic $o(n^2)$ -query $(3 - \epsilon)$ -approximation algorithms for

*A preliminary version of this paper appears in *Proceedings of the 30th Workshop on Combinatorial Mathematics and Computation Theory*, Hualien, Taiwan, April 2013, pp. 65–68.

[†]Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan. Email: clchang@saturn.yzu.edu.tw

[‡]Supported in part by the National Science Council of Taiwan under grant 101-2221-E-155-015-MY2.

any constant $\epsilon > 0$ [2]. This paper improves upon his result by showing that METRIC 1-MEDIAN has no deterministic $o(n^2)$ -query $(4 - \epsilon)$ -approximation algorithms for any constant $\epsilon > 0$.

In social network analysis, the importance of an actor in a network may be quantified by several centrality measures, among which the closeness centrality of an actor is defined to be its average distance to other actors [13]. So METRIC 1-MEDIAN can be interpreted as the problem of finding the most important point in a metric space. Goldreich and Ron [6] and Eppstein and Wang [5] present randomized algorithms for approximating the closeness centralities of vertices in undirected graphs.

2 Definitions

For $n \in \mathbb{N}$, denote $[n] \equiv \{1, 2, \dots, n\}$. Trivially, $[0] = \emptyset$. An n -point metric space $([n], d)$ is the set $[n]$, called the groundset, endowed with a function $d: [n] \times [n] \rightarrow \mathbb{R}$ satisfying

- (1) $d(x, y) \geq 0$ (non-negativeness),
- (2) $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles),
- (3) $d(x, y) = d(y, x)$ (symmetry), and
- (4) $d(x, y) + d(x, z) \geq d(y, z)$ (triangle inequality)

for all $x, y, z \in [n]$. An equivalent definition requires the triangle inequality only for distinct $x, y, z \in [n]$, axioms (1)–(3) remaining.

An algorithm with oracle access to a metric space $([n], d)$ is given n and may query d on any $(x, y) \in [n] \times [n]$ to obtain $d(x, y)$. Without loss of generality, we forbid queries for $d(x, x)$, which trivially return 0, as well as repeated queries, where a query for $d(x, y)$ is considered to repeat that for $d(y, x)$. For convenience, denote an algorithm ALG with oracle access to $([n], d)$ by ALG^d .

Given oracle access to a finite metric space $([n], d)$, the METRIC 1-MEDIAN problem asks for a point in $[n]$ with the minimum average distance to all points. An algorithm for this problem is α -approximate if it outputs a point $x \in [n]$ satisfying

$$\sum_{y \in [n]} d(x, y) \leq \alpha \min_{x' \in [n]} \sum_{y \in [n]} d(x', y),$$

where $\alpha \geq 1$.

The following theorem is due to Chang [3] and generalized by Wu [14].

Theorem 1 ([3, 14]). *METRIC 1-MEDIAN has a deterministic nonadaptive $O(n^{1.5})$ -time 4-approximation algorithm.*

3 Lower bound

Fix arbitrarily a deterministic $o(n^2)$ -query algorithm A for METRIC 1-MEDIAN and a constant $\delta \in (0, 0.1)$. By padding queries, we may assume the existence of a function $q: \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ such that A makes exactly $q(n) = o(n^2)$ queries given oracle access to any metric space with groundset $[n]$.

We introduce some notations concerning a function $d: [n] \times [n] \rightarrow \mathbb{R}$ to be determined later. For $i \in [q(n)]$, denote the i th query of A^d by $(x_i, y_i) \in [n] \times [n]$; in other words, the i th query of A^d asks for $d(x_i, y_i)$. Note that (x_i, y_i) depends only on $d(x_1, y_1), d(x_2, y_2), \dots, d(x_{i-1}, y_{i-1})$ because A is deterministic and has been fixed. For $x \in [n]$ and $i \in \{0, 1, \dots, q(n)\}$,

$$N_i(x) \stackrel{\text{def.}}{=} \{y \in [n] \mid \{(x, y), (y, x)\} \cap \{(x_j, y_j) \mid j \in [i]\} \neq \emptyset\}, \quad (1)$$

$$\alpha_i(x) \stackrel{\text{def.}}{=} |N_i(x)|, \quad (2)$$

following Chang [2] with a slight change in notation. Equivalently, $\alpha_i(x)$ is the degree of x in the undirected graph with vertex set $[n]$ and edge set $\{(x_j, y_j) \mid j \in [i]\}$. As $[0] = \emptyset$, $\alpha_0(x) = 0$ for $x \in [n]$. Note that $\alpha_i(\cdot)$ depends only on $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$. Denote the output of A^d by p . By adding at most $n - 1 = o(n^2)$ dummy queries, we may assume without loss of generality that

$$(p, y) \in \{(x_i, y_i) \mid i \in [q(n)]\} \quad (3)$$

for all $y \in [n] \setminus \{p\}$. Consequently,

$$\alpha_{q(n)}(p) = n - 1. \quad (4)$$

Fix any set $S \subseteq [n]$ of size $\lceil \delta n \rceil$, e.g., $S = \lceil \delta n \rceil$.

We proceed to construct d by gradually freezing distances. For brevity, freezing the value of $d(x, y)$ implicitly freezes $d(y, x)$ to the same value, where $x, y \in [n]$. Inductively, having answered the first $i - 1$ queries of A^d by freezing $d(x_1, y_1), d(x_2, y_2), \dots, d(x_{i-1}, y_{i-1})$, where $i \in [q(n)]$, answer the i th query by

$$d(x_i, y_i) = \begin{cases} 3, & \text{if } x_i, y_i \in S; \\ 3, & \text{if } x_i \in S, y_i \notin S \text{ and } \alpha_{i-1}(x_i) \leq \delta n; \\ 3, & \text{if } y_i \in S, x_i \notin S \text{ and } \alpha_{i-1}(y_i) \leq \delta n; \\ 4, & \text{if } x_i \in S, y_i \notin S \text{ and } \alpha_{i-1}(x_i) > \delta n; \\ 4, & \text{if } y_i \in S, x_i \notin S \text{ and } \alpha_{i-1}(y_i) > \delta n; \\ 2, & \text{if } x_i, y_i \notin S \text{ and } \max\{\alpha_{i-1}(x_i), \alpha_{i-1}(y_i)\} \leq \delta n; \\ 4, & \text{if } x_i, y_i \notin S \text{ and } \max\{\alpha_{i-1}(x_i), \alpha_{i-1}(y_i)\} > \delta n. \end{cases} \quad (5)$$

It is not hard to verify that the seven cases in equation (5) are exhaustive and mutually exclusive. We have now frozen $d(x_i, y_i)$ for all $i \in [q(n)]$ and none of the other distances. As repeated queries are forbidden, equation (5) does not freeze one distance twice, preventing inconsistency.

Set

$$B \stackrel{\text{def.}}{=} \{x \in [n] \mid \alpha_{q(n)}(x) > \delta n\}, \quad (6)$$

$$\hat{p} \stackrel{\text{def.}}{=} \operatorname{argmin}_{x \in S} \alpha_{q(n)}(x), \quad (7)$$

breaking ties arbitrarily. For all distinct $x, y \in [n]$ with $(x, y), (y, x) \notin \{(x_i, y_i) \mid i \in [q(n)]\}$, let

$$d(x, y) = \begin{cases} 1, & \text{if } x = \hat{p}, y \notin S \cup B; \\ 1, & \text{if } y = \hat{p}, x \notin S \cup B; \\ 3, & \text{if } x, y \in S \cup B; \\ 4, & \text{if } x \in (S \cup B) \setminus \{\hat{p}\} \text{ and } y \notin (S \cup B \cup \{\hat{p}\}); \\ 4, & \text{if } y \in (S \cup B) \setminus \{\hat{p}\} \text{ and } x \notin (S \cup B \cup \{\hat{p}\}); \\ 2, & \text{otherwise.} \end{cases} \quad (8)$$

Clearly, the six cases in equation (8) are exhaustive and mutually exclusive. Furthermore, equation (8) assigns the same value to $d(x, y)$ and $d(y, x)$. Finally, for all $x \in [n]$,

$$d(x, x) = 0. \quad (9)$$

Equations (5), (8) and (9) complete the construction of d by freezing all distances.

The following lemma is straightforward.

Lemma 2. *For all distinct $x, y \in [n]$, $d(x, y) \in \{1, 2, 3, 4\}$.*

Below is an immediate consequence of equation (7).

Lemma 3. $\hat{p} \in S$.

The following lemma is a consequence of equations (1)–(2) and our forbidding repeated queries.

Lemma 4. *For all $x \in [n]$ and $i \in [q(n)]$,*

$$\alpha_i(x) - \alpha_{i-1}(x) = \begin{cases} 0, & \text{if } x \notin \{x_i, y_i\}; \\ 1, & \text{otherwise.} \end{cases}$$

Proof. The case of $x \notin \{x_i, y_i\}$ is immediate from equations (1)–(2). Suppose that $x \in \{x_i, y_i\}$. By symmetry, we may assume $x = x_i$. So by equation (1),

$$N_i(x) = N_{i-1}(x) \cup \{y_i\}. \quad (10)$$

As $(x, y_i) = (x_i, y_i)$ is the i th query and we forbid repeated queries,

$$y_i \notin N_{i-1}(x) \quad (11)$$

by equation (1).¹ Equations (2) and (10)–(11) complete the proof. \square

¹In detail, if $y_i \in N_{i-1}(x)$, then $(x_j, y_j) \in \{(x, y_i), (y_i, x)\}$ for some $j \in [i-1]$ by equation (1); hence the i th query $(x_i, y_i) = (x, y_i)$ repeats the j th query, a contradiction.

In short, Lemma 4 says that adding the edge (x_i, y_i) to an undirected graph without that edge increases the degree of x by 1 if and only if $x \in \{x_i, y_i\}$.

Lemma 5. *For all $x \in [n]$ and $i \in [q(n) + 1]$, if $\alpha_{i-1}(x) > \delta n$, then $x \in B$.*

Proof. By Lemma 4, $\alpha_{q(n)}(x) \geq \alpha_{i-1}(x)$. Invoking equation (6) then completes the proof. \square

Lemma 6.

$$\sum_{x \in [n]} \alpha_{q(n)}(x) = 2q(n).$$

Proof. Recall that the left-hand side is the sum of degrees in the undirected graph with vertex set $[n]$ and edge set $\{(x_i, y_i) \mid i \in [q(n)]\}$. As we forbid repeated queries, $|\{(x_i, y_i) \mid i \in [q(n)]\}| = q(n)$. Finally, it is a basic fact in graph theory that the sum of degrees in an undirected graph equals twice the number of edges. \square

Lemma 7 (Implicit in [2, Lemma 13]). $|B| = o(n)$.

Proof. We have

$$|B| \delta n \stackrel{\text{equation (6)}}{\leq} \sum_{x \in B} \alpha_{q(n)}(x) \leq \sum_{x \in [n]} \alpha_{q(n)}(x) \stackrel{\text{Lemma 6}}{=} 2q(n).$$

This gives $|B| = o(n)$ as $\delta \in (0, 0.1)$ is a constant and $q(n) = o(n^2)$. \square

Lemma 8. *For all sufficiently large n and all $i \in [q(n) + 1]$,*

$$\alpha_{i-1}(\hat{p}) \leq \delta n. \tag{12}$$

Proof. By Lemma 7, $|S| = \lceil \delta n \rceil$ and $\delta \in (0, 0.1)$ being a constant, $S \setminus B \neq \emptyset$ for all sufficiently large n . By equation (6), $S \setminus B \neq \emptyset$ $\alpha_{q(n)}(x) \leq \delta n$ for some $x \in S$, which together with equation (7) gives $\alpha_{q(n)}(\hat{p}) \leq \delta n$. Finally, Lemma 4 and $\alpha_{q(n)}(\hat{p}) \leq \delta n$ imply inequality (12) for all $i \in [q(n) + 1]$. \square

Henceforth, assume n to be sufficiently large to satisfy inequality (12) for all $i \in [q(n) + 1]$.

Lemma 9. *For all $x, y \in [n]$, if $d(x, y) = 1$, then one of the following conditions is true:*

- $x = \hat{p}$ and $y \notin S \cup B$;
- $y = \hat{p}$ and $x \notin S \cup B$.

Proof. Inspect equation (8), which is the only equation that may set distances to 1. \square

Lemma 10. *For all distinct $x, y \in [n] \setminus (S \cup B)$, $d(x, y) = 2$.*

Proof. By Lemma 5, $\max\{\alpha_{i-1}(x_i), \alpha_{i-1}(y_i)\} > \delta n$ means $\{x_i, y_i\} \cap B \neq \emptyset$, where $i \in [q(n)+1]$. So only the second-to-last case in equation (5), which sets $d(x_i, y_i) = 2$, may be consistent with $x_i, y_i \notin S \cup B$.

By Lemma 3, $\hat{p} \in S$. So only the last case in equation (8), which sets $d(x, y) = 2$, may be consistent with $x, y \notin S \cup B$. \square

Lemma 11. *For all $x \in [n] \setminus \{\hat{p}\}$, $d(\hat{p}, x) \in \{1, 3\}$.*

Proof. By Lemma 3 and inequality (12), only the first three cases in equation (5), which set $d(x_i, y_i) = 3$, may be consistent with $x_i = \hat{p}$ or $y_i = \hat{p}$.

Again by Lemma 3, only the first three cases in equation (8), which set $d(x, y) \in \{1, 3\}$, may be consistent with $x = \hat{p}$ or $y = \hat{p}$. \square

Lemma 12. *There do not exist distinct $x, y, z \in [n]$ with $d(x, y) = 1$ and $\{d(x, z), d(y, z)\} = \{2, 4\}$.*

Proof. By Lemma 9, $d(x, y) = 1$ implies $\hat{p} \in \{x, y\}$. By symmetry, assume $x = \hat{p}$. Then $d(x, z) \in \{1, 3\}$ by Lemma 11. \square

Lemma 13. *There do not exist distinct $x, y, z \in [n]$ with $d(x, y) = d(x, z) = 1$ and $d(y, z) \in \{3, 4\}$.*

Proof. By Lemma 9, $d(x, y) = d(x, z) = 1$ implies $x = \hat{p}$ and $y, z \notin S \cup B$. Then $d(y, z) = 2$ by Lemma 10. \square

Lemmas 12–13 forbid all possible violations of the triangle inequality, yielding the following lemma.

Lemma 14. *$([n], d)$ is a metric space.*

Proof. Lemmas 2 and 12–13 establish the triangle inequality for d . Furthermore, d is symmetric because (1) freezing $d(x, y)$ automatically freezes $d(y, x)$ to the same value, (2) forbidding repeated queries prevents equation (5) from assigning inconsistent values to one distance and (3) equation (8) is symmetric. All the other axioms for metrics are easy to verify. \square

Recall that p denotes the output of A^d . We proceed to compare $\sum_{x \in [n]} d(p, x)$ with $\sum_{x \in [n]} d(\hat{p}, x)$.

Lemma 15. *There exist $k(1), k(2), \dots, k(n-1) \in [q(n)]$ and distinct $z_{k(1)}, z_{k(2)}, \dots, z_{k(n-1)} \in [n]$ such that*

$$\alpha_{k(t)-1}(p) = t - 1, \quad (13)$$

$$\alpha_{k(t)}(p) = t, \quad (14)$$

$$(p, z_{k(t)}) \in \{(x_{k(t)}, y_{k(t)}), (y_{k(t)}, x_{k(t)})\} \quad (15)$$

for all $t \in [n-1]$.

Proof. By Lemma 4, equation (4) and the easy fact that $\alpha_0(p) = 0$, there exist distinct $k(1), k(2), \dots, k(n-1) \in [q(n)]$ satisfying equations (13)–(14) for all $t \in [n-1]$.² Lemma 4 and equations (13)–(14) imply $p \in \{x_{k(t)}, y_{k(t)}\}$, establishing the existence of $z_{k(t)}$ satisfying equation (15). If $z_{k(1)}, z_{k(2)}, \dots, z_{k(n-1)}$ are not distinct, then there are repeated queries by equation (15), a contradiction. \square

From now on, let $k(1), k(2), \dots, k(n-1) \in [q(n)]$ and distinct $z_{k(1)}, z_{k(2)}, \dots, z_{k(n-1)} \in [n]$ satisfy equations (13)–(15) for all $t \in [n-1]$.

Lemma 16. *For each $t \in [n-1]$, if $t \geq \lceil \delta n \rceil + 2$ and $z_{k(t)} \notin S$, then $d(p, z_{k(t)}) = 4$.*

Proof. Assume in equation (15) that $p = x_{k(t)}$ and $z_{k(t)} = y_{k(t)}$; the other case will be symmetric. By equation (13),

$$\alpha_{k(t)-1}(x_{k(t)}) = t - 1 > \delta n. \quad (16)$$

Case 1: $x_{k(t)} \in S$. By equation (5), $x_{k(t)} \in S$ and $y_{k(t)} = z_{k(t)} \notin S$,

$$d(x_{k(t)}, y_{k(t)}) = \begin{cases} 3, & \text{if } \alpha_{k(t)-1}(x_{k(t)}) \leq \delta n; \\ 4, & \text{if } \alpha_{k(t)-1}(x_{k(t)}) > \delta n. \end{cases} \quad (17)$$

Case 2: $x_{k(t)} \notin S$. By equation (5), $x_{k(t)} \notin S$ and $y_{k(t)} = z_{k(t)} \notin S$,

$$d(x_{k(t)}, y_{k(t)}) = \begin{cases} 2, & \text{if } \max\{\alpha_{k(t)-1}(x_{k(t)}), \alpha_{k(t)-1}(y_{k(t)})\} \leq \delta n; \\ 4, & \text{if } \max\{\alpha_{k(t)-1}(x_{k(t)}), \alpha_{k(t)-1}(y_{k(t)})\} > \delta n. \end{cases} \quad (18)$$

Equation (16) together with any one of equations (17)–(18) implies $d(x_{k(t)}, y_{k(t)}) = 4$. Hence $d(p, z_{k(t)}) = d(x_{k(t)}, y_{k(t)}) = 4$. \square

We are now able to analyze the quality of p as a solution to METRIC 1-MEDIAN.

Lemma 17.

$$\sum_{x \in [n]} d(p, x) \geq 4(n - 2\lceil \delta n \rceil - 2).$$

Proof. By the distinctness of $z_{k(1)}, z_{k(2)}, \dots, z_{k(n-1)}$ in Lemma 15,

$$\sum_{x \in [n]} d(p, x) \geq \sum_{t \in [n-1]} d(p, z_{k(t)}). \quad (19)$$

Write $A = \{t \in [n-1] \mid z_{k(t)} \in S\}$. As $z_{k(1)}, z_{k(2)}, \dots, z_{k(n-1)}$ are distinct,

$$|A| \leq |S|. \quad (20)$$

²Observe that $\alpha_i(p)$ must go through all of $0, 1, \dots, n-1$ as i increases from 0 to $q(n)$.

Furthermore,

$$\begin{aligned}
& \sum_{t \in [n-1]} d(p, z_{k(t)}) \\
& \geq \sum_{t \in [n-1], t \geq \lceil \delta n \rceil + 2, t \notin A} d(p, z_{k(t)}) \\
& \stackrel{\text{Lemma 16}}{=} \sum_{t \in [n-1], t \geq \lceil \delta n \rceil + 2, t \notin A} 4 \\
& \geq 4(n - \lceil \delta n \rceil - 2 - |A|). \tag{21}
\end{aligned}$$

Equations (19)–(21) and $|S| = \lceil \delta n \rceil$ complete the proof. \square

We now analyze the quality of \hat{p} as a solution to METRIC 1-MEDIAN. The following lemma is immediate from equation (8).

Lemma 18. *For all $y \in [n] \setminus (S \cup B)$, if $y \neq \hat{p}$ and $(\hat{p}, y), (y, \hat{p}) \notin \{(x_j, y_j) \mid j \in [q(n)]\}$, then $d(\hat{p}, y) = 1$.*

Lemma 19.

$$\sum_{y \in [n]} d(\hat{p}, y) \leq n + 3 \cdot (\lceil \delta n \rceil + o(n) + \delta n).$$

Proof. By equation (1),

$$N_{q(n)}(\hat{p}) = \{y \in [n] \mid \{(\hat{p}, y), (y, \hat{p})\} \cap \{(x_j, y_j) \mid j \in [q(n)]\} \neq \emptyset\}.$$

This and Lemma 18 imply $d(\hat{p}, y) = 1$ for all $y \in [n] \setminus (S \cup B)$ with $y \neq \hat{p}$ and $y \notin N_{q(n)}(\hat{p})$. Therefore,

$$\sum_{y \in [n] \setminus (S \cup B \cup N_{q(n)}(\hat{p}))} d(\hat{p}, y) \leq n - |S \cup B \cup N_{q(n)}(\hat{p})|. \tag{22}$$

Clearly,

$$\sum_{y \in S \cup B \cup N_{q(n)}(\hat{p})} d(\hat{p}, y) \stackrel{\text{Lemma 2}}{\leq} \sum_{y \in S \cup B \cup N_{q(n)}(\hat{p})} 4 = 4 \cdot |S \cup B \cup N_{q(n)}(\hat{p})| \tag{23}$$

Furthermore,

$$|N_{q(n)}(\hat{p})| \stackrel{\text{equation (2)}}{=} \alpha_{q(n)}(\hat{p}) \stackrel{\text{inequality (12)}}{\leq} \delta n.$$

This and Lemma 7 imply

$$|S \cup B \cup N_{q(n)}(\hat{p})| \leq \lceil \delta n \rceil + o(n) + \delta n \tag{24}$$

as $|S| = \lceil \delta n \rceil$. To complete the proof, sum up inequalities (22)–(23) and then use inequality (24) in the trivial way. \square

Combining Lemmas 14, 17 and 19 yields our main theorem, stated below.

Theorem 20. METRIC 1-MEDIAN has no deterministic $o(n^2)$ -query $(4 - \epsilon)$ -approximation algorithm for any constant $\epsilon > 0$.

Proof. Lemma 14 asserts that $([n], d)$ is a metric space. By Lemmas 17 and 19,

$$\sum_{x \in [n]} d(p, x) \geq 4(1 - 8\delta - o(1)) \sum_{x \in [n]} d(\hat{p}, x).$$

This proves the theorem because the deterministic $o(n^2)$ -query algorithm A and the constant $\delta \in (0, 0.1)$ are picked arbitrarily (note that p denotes the output of A^d). \square

Theorem 20 complements Theorem 1.

It is possible to simplify equation (8) at the expense of an additional assumption. Without loss of generality, we may assume that $\alpha_{q(n)}(x) = n - 1$ for all $x \in B$; this increases the query complexity by a multiplicative factor of $O(1)$ by equation (6). Therefore, if $x \in B$ or $y \in B$, then $d(x, y)$ will be frozen by equation (5). So the third to fifth cases in equation (8), which satisfies $x \in B$ or $y \in B$, can be omitted.

References

- [1] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k -median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.
- [2] C.-L. Chang. Some results on approximate 1-median selection in metric spaces. *Theoretical Computer Science*, 426:1–12, 2012.
- [3] C.-L. Chang. Deterministic sublinear-time approximations for metric 1-median selection. *Information Processing Letters*, 113(8):288–292, 2013.
- [4] K. Chen. On coresets for k -median and k -means clustering in metric and Euclidean spaces and their applications. *SIAM Journal on Computing*, 39(3):923–947, 2009.
- [5] D. Eppstein and J. Wang. Fast approximation of centrality. *Journal of Graph Algorithms and Applications*, 8(1):39–45, 2004.
- [6] O. Goldreich and D. Ron. Approximating average parameters of graphs. *Random Structures & Algorithms*, 32(4):473–493, 2008.
- [7] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O’Callaghan. Clustering data streams: Theory and practice. *IEEE Transactions on Knowledge and Data Engineering*, 15(3):515–528, 2003.

- [8] P. Indyk. Sublinear time algorithms for metric space problems. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 428–434, 1999.
- [9] P. Indyk. *High-dimensional computational geometry*. PhD thesis, Stanford University, 2000.
- [10] R. Jaiswal, A. Kumar, and S. Sen. A simple D^2 -sampling based PTAS for k -means and other clustering problems. In *Proceedings of the 18th Annual International Conference on Computing and Combinatorics*, pages 13–24, 2012.
- [11] A. Kumar, Y. Sabharwal, and S. Sen. Linear-time approximation schemes for clustering problems in any dimensions. *Journal of the ACM*, 57(2):5, 2010.
- [12] R. R. Mettu and C. G. Plaxton. Optimal time bounds for approximate clustering. *Machine Learning*, 56(1–3):35–60, 2004.
- [13] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [14] B.-Y. Wu. On approximating metric 1-median in sublinear time. *Information Processing Letters*, 114(4):163–166, 2014.